

Chapter 32

Medical Information Extraction in European Portuguese

Liliana Ferreira

University of Aveiro, Portugal

António Teixeira

University of Aveiro, Portugal

João Paulo Silva Cunha

University of Aveiro, Portugal

ABSTRACT

The electronic storage of medical patient data is becoming a daily experience in most of the practices and hospitals worldwide. However, much of the available data is in free text form, a convenient way of expressing concepts and events but especially challenging if one wants to perform automatic searches, summarization, or statistical analyses. Information Extraction can relieve some of these problems by offering a semantically informed interpretation and abstraction of the texts. MedInX, the Medical Information eXtraction system presented in this chapter is designed to process textual clinical discharge records in order to perform automatic and accurate mapping of free text reports onto a structured representation. MedInX components are based on Natural Language Processing principles and provide several mechanisms to read, process, and utilize external resources, such as terminologies and ontologies. MedInX current practical applications include automatic code assignment and an audit system capable of systematically analyze the content and completeness of the clinical reports. Recent evaluation efforts on a set of authentic patient discharge letters indicate that the system performs with 95% precision and recall.

INTRODUCTION

The rapid adoption of Electronic Health Records (EHR) in the clinical domain, with the parallel growth of narrative data in electronic form, is a strong incentive for the development of Medical

Language Processing (MLP) systems. However, much of the available clinical data is in narrative form. While narratives are a convenient way of expressing concepts and events, they can be difficult if one wants to perform automatic searches, summarization, decision support or statistical

DOI: 10.4018/978-1-4666-3986-7.ch032

analysis. To overcome this problem and also to ensure improved quality control and reduced medical errors, structured data is required. This is where Natural Language Processing (NLP) and more precisely Information Extraction (IE) is needed.

Electronic access to health information is becoming more frequent worldwide through numerous electronic health information systems, which generate millions of gigabytes of health information annually, more than in many other domains (Cios and Moore, 2002). In fact, this type of access to information is critical for improved health care and to reduce medical errors. For example, in Portugal, with a population of approximately ten million inhabitants, the latest available statistics (National Institute of Statistics, 2009) report that every year there is almost a million of hospital admissions and forty six million specialized care outpatient visits in the National Health System.

Several efforts are currently being conducted to generalize the use of EMRs in the Portuguese health institutions, but most of the information available in these systems is in textual form and, even if electronically available, remains locked up within text. Enriching EMR systems with rich domain knowledge and rules would greatly enhance their performance and ability to support clinical decisions. However, the content of health records is extremely complex, as a considerable part of a patient health records is free-form written or spoken text. Particularly, they describe a sequence of events, narratives, reflecting the need for a precise and complete explanation when describing the health status of a patient. This type of expressive description also bears a substantial ambiguity and personal differences in vocabulary and style (Lovis et al., 2000; Suominen, 2009). Frequently, specialists from the same discipline cannot agree on unambiguous terms to be used while describing a patient's condition.

Information search from this type of narrative text is difficult and time consuming. Human Language Technology (HLT) and, particularly, NLP,

is increasingly gaining the interest of both health care practitioners and academic researchers by offering the possibility of extracting precise facts from a document set and of finding interesting associations among disparate facts, leading to the discovery of new or unsuspected knowledge (Ananiadou and Mcnaught, 2005) instead of leaving the health professional with the problem of having to read several tens of thousands of documents.

The present chapter focuses on MedInX, a *Medical Information eXtraction* system tailored to process textual clinical discharge records in order to perform automatic and accurate mapping of free text reports onto a structured representation. MedInX uses IE technology to structure the information present in discharge reports originated by the EHR system being used in the region of Aveiro in Portugal, the Telematic Healthcare Network RTS® (Cunha et al., 2006), and automatically instantiates a knowledge representation model from the free-text patient discharge letters.

The main goal of MedInX is to improve access to clinical reports, and, consequently, enable faster and more accurate statistical data creation and analysis. Improving the access to quality information on health for both patient and health professionals also contributes to reduce medical errors and to increase safety and efficiency, while the creation of a semantically informed interpretation of texts contributes to enable patient mobility and borderless access to health care. Moreover, MedInX reaches the previous objectives while allowing physicians to continue to practice using their current descriptive language in free-text reports without a requirement to enter structured data in a complex, time-consuming computer-based system.

The current chapter starts by introducing some broad definitions related with Medical Information Extraction and Knowledge Representation while reviewing some literature published in the areas. The chapter continues by presenting the main challenges of Medical Information Extraction in European Portuguese and the Telematic Health-

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/medical-information-extraction-european-portuguese/77165

Related Content

Using Software to Deliver Language Intervention in Inclusionary Settings

Mary Sweig Wilson and Jeffrey Pascoe (2010). *Handbook of Research on Human Cognition and Assistive Technology: Design, Accessibility and Transdisciplinary Perspectives* (pp. 132-156).

www.irma-international.org/chapter/using-software-deliver-language-intervention/42833

Putting the Content Into Context: Features and Gaps in Image Retrieval

Henning Müller and Jayashree Kalpathy-Cramer (2009). *International Journal of Healthcare Information Systems and Informatics* (pp. 88-98).

www.irma-international.org/article/putting-content-into-context/2244

Changing Healthcare Institutions with Large Information Technology Projects

Matthew W. Guah (2010). *Health Information Systems: Concepts, Methodologies, Tools, and Applications* (pp. 1304-1317).

www.irma-international.org/chapter/changing-healthcare-institutions-large-information/49932

Improving Access to Oncology Care for Individuals and Families through Telehealth

Johanna den Duyf, Lars Aplan, Valerie Ashworth, Arminee Kazanjian, Margarita Loyola and Sarah Robertson (2010). *Handbook of Research on Developments in E-Health and Telemedicine: Technological and Social Perspectives* (pp. 669-688).

www.irma-international.org/chapter/improving-access-oncology-care-individuals/40671

Emerging ICT Challenges on Provision of Online HIV/AIDS Advisory Services

Simon Samwel Msanjila (2013). *Handbook of Research on ICTs for Human-Centered Healthcare and Social Care Services* (pp. 248-269).

www.irma-international.org/chapter/emerging-ict-challenges-provision-online/77146